

**Universidad de Buenos Aires**  
**Facultad de Ciencias Exactas y Naturales**  
**Facultad de Ingeniería**  
**Maestría en Explotación de Datos y Descubrimiento de Conocimiento**  
**Aprendizaje Automático**  
**1er cuatrimestre de 2021**  
**Trabajo práctico Nro 2**

El objetivo de este trabajo práctico es analizar las particularidades de la utilización de algoritmos de ensambles aplicados en casos casi reales. El mismo pretende fijar conceptos estudiados en la teoría: métodos de ensamble; random forests; boosting; sobreajuste; tolerancia al ruido. El material básico para la elaboración del presente trabajo se encuentra en las teóricas y prácticas presentadas hasta el momento y en las próximas clases y en las referencias bibliográficas indicadas [1,2,3]. Podrá utilizarse cualquier otra fuente siempre que esté correctamente referenciada.

El presente trabajo será grupal. El grupo deberá estar compuesto por exactamente tres integrantes. Se evaluarán los contenidos del Trabajo Práctico durante el coloquio posterior a la entrega del TP junto al material teórico y práctico enseñado en las clases de toda la materia. Todos los integrantes deben tener conocimiento del desarrollo del TP.

**La fecha límite de entrega es el lunes 28 de junio a las 17:00 hs.**

Para el desarrollo del trabajo se utilizará el conjunto de datos Ryerson Audiovisual Database of Emotional Speech and Song (RAVDESS) <https://zenodo.org/record/1188976#.YLbG6aJKiV4> con el objetivo de predecir a partir del habla, qué emociones se expresaron.

Se deberá elaborar un informe preferentemente en LaTeX y entregarlo en formato .pdf. La entrega deberá estar acompañada de la Jupyter Notebook en Python utilizada para generar los resultados. El documento a entregar debe cumplir con los siguientes requisitos:

- debe tener no más que cuatro hojas, con fuente tamaño 10 e interlineado simple. La bibliografía no cuenta en la cantidad de hojas.
- una carátula en donde figuren universidad, nombre de maestría, materia, número de grupo, nombres de los integrantes del grupo, número de TP, año de cursada, etc. La carátula no cuenta en la cantidad de hojas.
- un resumen (del estilo de un artículo científico de no más de 200 palabras)
- una introducción en donde, entre otros, conste el objetivo del trabajo y una explicación de cómo está organizado el resto del documento.
- una sección de datos, en donde se describan los datos utilizados y sus particularidades

- una sección de metodología, en donde se describan las metodologías utilizadas (sobre datos y sobre algoritmos)
- una sección resultados, que incluya los resultados y su análisis
- una sección de conclusiones. Por tratarse de un trabajo de investigación netamente práctico, las conclusiones deben ser la resultante de la elaboración de las pruebas realizadas. La información obtenida de referencias externas puede y debe ser tomada como insumo, pero no como conclusión.
- referencias bibliográficas (referenciadas a lo largo del trabajo)

El informe se deberá publicar en el aula virtual de la materia por uno sólo de los integrantes del grupo.

Para realizar el informe se deberán considerar y documentar los siguientes puntos:

- a) Obtener el dataset (modalidad audio) a partir del enlace provisto. Notar que los zips que corresponden a solo audio son: [Audio\\_Song\\_Actors\\_01-24.zip](#) y [Audio\\_Speech\\_Actors\\_01-24.zip](#). Una vez descargado, extraer el conjunto de atributos [eGeMAPS](#) utilizando la librería [opensmile-python](#).
- b) Elegir un modelo de ensamble visto durante las clases y entrenarlo para predecir emociones dividiendo el dataset de las siguientes maneras:
  - i) 12-fold cross validation armando los folds de forma aleatoria.
  - ii) Leave-2-speakers out. Es decir, 12 folds conteniendo cada uno 2 actores distintos. Se recomienda utilizar la función de sklearn [LeaveOneGroupOut](#)

Reportar los resultados obtenidos y realizar un análisis de los mismos, teniendo en cuenta las características del dataset utilizado.

- c) Realizar la división de datos de la manera que considere más adecuada, teniendo en cuenta el análisis realizado en b), y comparar el desempeño de Random Forests y AdaBoost. Realizar una búsqueda de hiperparámetros para ambos modelos. Para esto, deberá elegirse una métrica de evaluación y justificar la elección. Reportar y analizar la matriz de confusión del mejor modelo, tanto para el caso de audios cantados como hablados.
- d) Elegir y realizar alguno de los siguientes puntos:
  - i) Evaluar la robustez del mejor modelo ante la presencia de ruido en los audios. Realizar una curva que muestre cómo varía una métrica de desempeño para distintas relaciones señal a ruido. El ruido utilizado puede ser sintético (gaussiano por ejemplo), o ruido ambiental. Algunas fuentes de ruido ambiental son:

<https://github.com/qutsaivt/QUT-NOISE>

<https://www.kaggle.com/aanhari/demand-dataset>

<https://urbansounddataset.weebly.com/urbansound8k.html>

- ii) Evaluar el mejor modelo en audios grabados por los integrantes del grupo. Pueden evaluar la generalización ante distintas frases, pronunciaciones, intensidad de la emoción, etc...
- iii) Calcular la media y desvío estándar de los atributos para cada actor, y normalizar los audios de cada actor utilizando esas estadísticas. Evaluar el desempeño del mejor modelo cuando es entrenado con los datos normalizados de esta manera (por hablante). Compare el desempeño con el modelo sin normalización. Explicar cuáles son las ventajas y desventajas de aplicar este preprocesamiento.

Opcionales (dan puntos extra):

- 1) Sumar al punto c) un modelo de perceptrón multicapa.
- 2) Evaluar el desempeño de los modelos al utilizar el conjunto de atributos [ComParE 2016](#).

### Referencias

- [1] An Introduction to Statistical Learning. Capítulos 2 (2.2.2), 5 (5.2) y 8 (8.2, 8.3.3, 8.3.4). <http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf>
- [2] Seni, Elder, "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions", Morgan & Claypool, 2010. [https://doc.lagout.org/Others/Data Mining/Ensemble Methods in Data Mining\\_ Improving Accuracy through Combining Predictions %5BSeni %26 Elder 2010-02-24%5D.pdf](https://doc.lagout.org/Others/Data Mining/Ensemble Methods in Data Mining_ Improving Accuracy through Combining Predictions %5BSeni %26 Elder 2010-02-24%5D.pdf)
- [3] <http://scott.fortmann-roe.com/docs/BiasVariance.html>

### Material suplementario:

Ejemplo de paper del tema: <https://arxiv.org/abs/1912.10458>

Teoría de los atributos de habla:

■ (Springer Theses) Florian Eyben (auth.) - Real-time Speech and Music Classification by L...